CHAPTER 6

QUALITY CONTROL

During the 1992 Air Force Health Study (AFHS) followup, stringent adherence to quality assurance (QA) was planned for and upheld throughout the study, from project initiation to final product delivery and acceptance by the Air Force. This chapter provides an overview of the specific QA measures developed and used by the project team, specifically in the areas of questionnaire and physical examination quality control (QC), laboratory QC measures, data management QC, statistical QC, and administrative QA.

QUESTIONNAIRE QUALITY CONTROL

The National Opinion Research Center (NORC) used both onsite and home office procedures to produce a comprehensive, high-quality data set. All AFHS questionnaires were pretested to evaluate completion time and participant acceptability before they were used during the 1992 followup. Onsite QC procedures included observing and rating interviewers, and reviewing every questionnaire twice at the completion of the interview (once by the interviewer who administered the questionnaire and then again by NORC's onsite supervisor or editor). Science Applications International Corporation (SAIC) conducted a review of 10 percent of the questionnaires for management acceptance throughout the program. SAIC reviewed 100 percent of the questionnaires during the first 4 weeks of the physical examination process. The Air Force also continuously conducted QA observations of all onsite activities.

QC of data processing included the following:

- Manual editing of each questionnaire
- 100-percent blind verification of data entry by a second key entry operator
- Computerized data cleaning to identify values out of range, inconsistent responses, and logic and arithmetic errors
- Review of response frequencies
- Review of the actual questionnaires to reconcile or correct detected errors.

NORC recruited and trained 11 interviewers according to the procedures described in Chapter 3. A minimum number of interviewers was selected to reduce variability between interviewing techniques. Additionally, the interviewers were blind to the participants' exposure status to avoid bias. Interviewers were required to ask questions exactly as written, and in the order in which they appeared. No personal interpretation was allowed.

NORC's onsite supervisor closely monitored both the staff of interviewers and the onsite editor. The supervisor reported to NORC's Project Manager at least weekly, and she was in turn evaluated by the Project Manager at the beginning of the study and during quarterly site visits.

Interviewers were closely observed in training to ensure that they were able to administer the questionnaire and record responses smoothly and correctly. During the study, the onsite supervisor checked interviewers for accuracy in following questionnaire skip patterns, circling the correct codes, and controlling the interview, including voice quality, reading, and use of associated forms and documents. The supervisor observed at least one interview per interviewer per quarter, and gave immediate retraining if she observed any errors.

Either the supervisor or NORC's onsite editor reviewed and edited each questionnaire immediately following each interview, and reported any errors to the supervisor, who retrained interviewers during daily contacts. Generalizations from individual interviews were used to train the entire group of interviewers. Whenever possible, missing information was retrieved from participants before they left the examination site. If errors were discovered when participants were no longer onsite, information was retrieved from them by telephone.

Once participant questionnaires were received by NORC's home office for data processing, they were reviewed for completeness by a coding supervisor and staff. The coding staff resolved inconsistencies in the questionnaires and prepared the forms for data entry. This included coding of open-ended responses, such as in the category "occupation." Ten percent of open-ended items for each batch of questionnaires were recoded. When a batch failed the 10-percent recode, the entire batch was recoded and the coding staff was retrained.

Key entry of data was 100-percent blind, verified by a second key entry operator. Interval Questionnaire data were passed through a computer program that checked for out-of-range data, inter-item inconsistencies, and logic and arithmetic errors. When discrepancies were detected, the questionnaires were reviewed and the errors corrected. Response frequencies also were reviewed, and any anomalies or errors previously undetected were corrected by reviewing the questionnaires. The process continued until no errors were found. All corrections were documented and entered into the data base, but no changes were made to the original data recorded in the questionnaires.

Baseline Questionnaire data was subject to reviews of response frequencies and cross-tabulations of related variables. Again, corrections were documented and entered into the data base, but the original data recorded in the questionnaires were not altered.

Diet Assessment forms were coded, read into electronic form, and cleaned by the subcontractor, Willett Associates. NORC performed a 10-percent check of the data delivered by Willett, consisting of an item-by-item comparison of answers recorded on the hard copy and those contained in the data base. Diet Assessment forms were checked in batches. No batches failed the QA check in the 1992 study. If any batch had failed the QA step, that batch would have been returned to the subcontractor for recoding, re-entry and recleaning. NORC performed a final recoding of Diet Assessment data to make missing values codes consistent with the rest of the Interval Questionnaire.

PHYSICAL EXAMINATION QUALITY CONTROL

QA was emphasized in administering the physical examination, as this data source provided a large part of the medical information for clinical and epidemiologic analyses.

Initial concern for a high-quality physical examination was addressed by a stringent Scripps Clinic and Research Foundation (SCRF) selection process for all personnel who were to interact directly with the participants. Each staff member was hand-selected for the AFHS on the basis of expertise, experience, and a commitment to remain with the study throughout the examination cycle. Furthermore, the Air Force reviewed the credentials of all key staff members and approved their participation in the study.

A complete pretest physical examination, interview, psychological test, and laboratory workup was done for 10 volunteers several weeks before the scheduled start of the study. The dermatologists received refresher training to enhance their skill in diagnosing chloracne, internists were provided with a review of techniques for detecting specific heart sounds, and diagnosticians were reminded to review Baseline, 1985, and 1987 examination data as they formulated all diagnoses. Furthermore, all aspects of patient contact were reviewed: the initial inbriefing of the participants, the logistics of transportation and patient flow within the clinic, and the final outbriefing by the diagnostician.

During the examinations, refinements continued whenever operational problems were detected by the SCRF staff and the Air Force onsite monitor, or when participants identified areas requiring improvement. Both of these types of information were addressed during the weekly clinical QA meeting of key SCRF staff. Written critique forms submitted by all participants also were reviewed in detail at the SCRF weekly meetings, providing additional insight to both temporary shortcomings of the entire logistic process and the numerous strong points of the programs.

Following examination of each participant group, all physical examination forms were reviewed by the SCRF staff for omissions, incomplete examinations, and inconsistencies. The examiners or technicians quickly were contacted to correct the data. Special effort was made to complete this review while the participants were at the examination site. In all cases of data correction, a complete audit trail was maintained. All mark-sense physical examination forms were read by an optical scanner as an ongoing QA of form completion. (This subject is discussed in more detail in the Data Management Quality Control section of this chapter.)

Compliance with all aspects of the physical examination process was monitored daily by the Air Force onsite monitor and the SCRF administrative team. Additional periodic inspections were conducted by the SCRF Chief of Medicine and the SAIC Project Manager. All such clinical reviews were done unobtrusively and with the full consent of the participant; suggestions or corrections to the examination procedure were always discussed privately with the attending physician. These inspections emphasized aspects of clinical techniques, sequence, and completeness of the clinical data with respect to the examination forms, and the blindness of the examinations. Of particular note were the detailed daily log entries of the five Air Force monitors. These entries ensured continuity of knowledge (the monitors rotated approximately every 2 weeks) by documenting examination procedural changes and recording events requiring followup by either the Air Force or SAIC.

Establishing a rapport with each study participant was a primary goal of all the organizations involved in this study. Although "rapport building" may not be a traditional QA parameter in most research studies, it is paramount in the AFHS because maintaining the satisfaction of participants encourages them to continue in the study, thus helping to avoid a significant reduction in future statistical power or introducing bias, or both. Therefore, every staff member, from the initial telephone recruiter to the nurse coordinator and the Project Manager, emphasized courtesy, empathy, assistance, and personalized treatment of each participant. Based on the evaluation forms, 73.5 percent of the participants evaluated their experience in the 1992 followup as excellent, and 22 percent classified it as good. Only 3.9 percent of the participants rated the experience as satisfactory, and only 0.7 percent felt that it was unsatisfactory.

LABORATORY QUALITY CONTROL

Before the study began, specific QC laboratory procedures were designed, developed, and implemented to rapidly detect problems related to test and assay performance, validity of reagents, analysis of data, and reporting of results. All laboratory assays for the study were performed with state-of-the-art laboratory equipment and techniques. Laboratory facilities all had the equivalent of National Institutes of Health (NIH) Biosafety Level 2 approval ratings and were certified by the College of American Pathology.

Quality Control Procedures for the Clinical Laboratory

The following list outlines the tests that were performed and the methods and equipment used:

- Hematology assays were performed on Coulter S-Plus® equipment.
- Sedimentation-rate determinations were performed using the large-tube Westergren method.
- Biochemical assays were performed using Baxter/Dade Paramax® Automated Chemical Analyzer.
- Radioimmunoassays were performed with standard test kits.
- Electrophoresis and occult blood tests were performed manually.
- Hepatitis B tests were performed using Abbott Diagnostic® kits.
- Monospecific antibodies were used for immunologlobulin assays using the Beckman Array Protein System®.
- Blood-cell counts were performed with standard microscopy.

- All urinalyses were performed using Clinitek®, a reflectance spectrometry urinalysis.
- All other assays were done using industry-standard equipment and techniques.

All laboratory operations were controlled with the use of an integrated medical laboratory management information system that incorporated direct device-to-data base interfaces for automated testing equipment. Data entry for manual tests was performed by the laboratory technologists. An automated audit trail and a set of comments for technologist remarks were kept for each test so that any QC results could be retraced.

Procedural QC included using the same instrument and reagents from the same lot numbers whenever possible throughout the study. If single lots were unavailable, an overlap analysis of both lots was used. Strict standards of calibration for all automated laboratory equipment were maintained at all times.

Trilevel or bilevel controls were used as the primary means for monitoring the quality of all tests. On every group of participant samples, one control (low, medium, or high) was run at the start, after every 9th sample, and at the end of each test run. Each trilevel control was used before repeating it in the run, when more than 18 experimental samples were analyzed. In addition, split aliquots were made from every 10th participant sample and were analyzed separately to measure test reproducibility. In radiommunoassays, all three control levels were run initially to validate the standard curve generated.

All QC data were analyzed and summarized in formal QC reports generated monthly. QC data were subjected to independent statistical analysis by the Air Force to produce and analyze time-dependent trends. For all equipment malfunctions or other exceptions, a formal QC exception report was prepared by the responsible individual and forwarded to the project management team.

An additional measure of QC used during the study was the cumulative sum (CUSUM) tests run with trilevel controls (1). In particular, the fast initial response (FIR) CUSUM QC technique was used in detecting long-term subtle drift that could have substantial adverse analytical consequences (2). FIR is a special case of the CUSUM QC scheme that increases the overall effectiveness of the QC procedure. Unlike QC procedures using standard control charts, which compare each observation to designated limits, these tests utilize the CUSUM of deviations from a target value.

CUSUM statistics were accumulated for each of the trilevels to quickly detect instrument calibration problems as identified by excessive drift. If an out-of-control situation was indicated, the graph showed when the change first occurred. When the CUSUM indicated an out-of-control situation, all adjacent patient samples were reanalyzed after the equipment was thoroughly checked and fresh controls were run. Coefficient of variation (CV) requirements were established for each test prior to the beginning of the physical examination process.

FIR CUSUM generally has been applied to QC in industry, particularly in high-volume, high-precision applications. It is believed that FIR CUSUM generally has not been applied in a biomedical setting, but it has proved to be effective in the AFHS.

As the examination portion of this study ended, laboratory outliers were analyzed for logical validity by an independent clinician. All out-of-range test results were examined and scored as clinically explainable, clinically possible, or clinically unexplained. No clinical laboratory data were excluded because all potential out-of-range results were found to be clinically explainable or clinically possible.

Quality Control Procedures for the Immunology Laboratory

The QC procedures for the cellular immunology section of the AFHS were structured to rapidly detect any problems in four major test parameters: assay performance, reagent validity, data analysis, and results reporting. The cellular immunology laboratory supervisor monitored compliance daily. Key aspects of the program included instrument and equipment calibration and maintenance, assay controls, accuracy and precision determination, and system failure checks.

The following QC measures were adhered to in all cellular immunology assays:

- Testing of a blood sample from a normal, healthy control individual with each group of AFHS patient samples.
- Duplicate testing of one random participant sample in each assay.
- Quadruplicate testing of each participant sample for each variable in each of the functional assays (e.g., phytohemagglutinin [PHA] stimulation, natural killer cell (NKC), and mixed lymphocyte culture).
- Parallel testing and monitoring reactivity of various lots of reagents when appropriate.
- Verification of participant and specimen identification by at least two individuals before final reporting to the data base.
- Note codes attached to any data point with a detected deviation due to procedural setup error, assay malfunction, equipment malfunction, or assay technical error.
- Note codes attached to any data point outside the range of expected values as identified by the cellular immunology laboratory supervisor.
- Review of all final assay reports by the cellular immunology laboratory supervisor prior to entry into the data base.

QC for each functional assay (including PHA, NKC, and mixed lymphocyte culture), consisted of monitoring assay controls, duplicate sample reproducibility, and trends in reagent reactivity. Assay precision was determined by calculating the CV of the quadruplicates for each variable tested. Also, a mean value of the CV for each assay was calculated. Individual CVs of 15 percent or less were the target values for the stimulated samples in the mitogen and NKC assays. The Student's t-test was applied to duplicates to determine if there was a significant difference in sampling for the functional assays. Critical t-values at the 0.05 significance level were used to determine if duplicate sample results varied significantly. Positive and negative values were assigned, arbitrarily subtracting the second duplicate value from the first, to determine if there was a systematic bias in one direction. Grubbs' statistical test (3) was used to identify any statistically significant outlier. This test was applied only to samples whose CVs were greater than 20 percent at a p-value of 0.01. The PHA stimulation effect was followed by daily evaluation of the radioactive counts in counts per minute. When counts fell below expected values, suggesting that reagent deterioration had occurred, new aliquots were used.

QC measures for the cell-surfaced marker assays included calculation of (CD4 + CD8)/CD3 (formerly $[T_4 + T_8]/T_{11}$) cell ratios, evaluation of flow cytometer computer outputs (cytograms and histograms), and duplicate sample testing. The cellular ratios should approximate the value 1.0 for a normal population. Validity of cytogram and histogram distributions generated by the flow cytometer was confirmed by the cellular immunology laboratory supervisor for each sample analyzed. The proportional difference between duplicate samples was calculated and monitored for significant differences.

DATA MANAGEMENT QUALITY CONTROL

Overview of Quality Control Procedures

The QC program for the data management activity consisted of multiple checks at all steps of the examination, data collection, and data processing cycle. Data QC procedures for data collection, conversion, and integration were developed before the clinical examinations began. Pretesting of all forms was conducted 4 weeks before the examinations actually began. Additionally, during the first 2 months of the clinical examinations, all data collection activities were intensely scrutinized to detect and correct procedural deficiencies. QC activities also included the following:

- Automated QC techniques applied to laboratory data
- Clinical evaluations of all laboratory outliers
- Review of all physical examination findings by one of two diagnosticians who was not involved in the conduct of the physical examinations
- Automated and manual data quality checking of hard copy against transcribed computer files for all questionnaire, physical examination, and medical coding data streams.

Five interwoven layers of QC were instituted to ensure data integrity. Efforts focused on data processing system design, design and administration of all exams or questionnaires, data completeness checks, data validation, and QC of medical records coding.

Data Processing System Design

Standards were established for data element formats (character or numeric), data element naming conventions, data element text labels, numeric codes for qualitative responses and results, QC range checks for continuous data elements, and QC validity checks for categorical data. A data dictionary provided detailed information on each data element.

A systems integration approach was applied to the design and implementation of data collection procedures so that data emanating from study sources (physical examination, questionnaire, laboratory) were consistent in file format and structure. This approach was necessary to ensure that all data could be integrated into a single data base management system for analysis. Figure 6-1 provides an overview of the QC activities used in the data management process.

Forms and questionnaires were carefully designed to ensure that all required data elements would be collected in accordance with the Study Protocol and in a standardized format. The design of these instruments was such that they reflected the order in which the examination itself would be administered and provided for the sequential recoding of information to streamline remaining data management activities.

Completed clinical examination forms and questionnaires were converted from hard copy to machine-readable images using customized data-entry systems or state-of-the-art optical mark reading equipment. Verification procedures were performed to ensure that a uniquely identified participant record existed within each data file, and that the appropriate number of responses for each applicable field was provided. Data files were then verified against original data sheets and corrected as necessary.

Data files were then subjected to validity checks. Any potentially conflicting results, as well as any data values falling at the extremes of expected ranges, were manually reviewed. Extreme values were reverified against the original raw data copies and either corrected or documented as valid results. Potentially conflicting results were returned to the examiners for review. These results were then documented as having been correctly recorded, corrected, or flagged for exclusion from analysis because of unresolvable examiner errors or omissions. This process was continued until all results were properly documented.

Once the edits were completed and the data reverified, the "cleaned" files or tapes were transferred to the data analysis center for final inspection and integration into the study data base. For this QC measure, each data file was loaded into a SAS® data set, and descriptive analyses were run. The validation, correction, transmission, and analysis QC procedures were repeated as necessary to ensure that all extreme or suspicious values had been validated.

Data QC Flow Chart

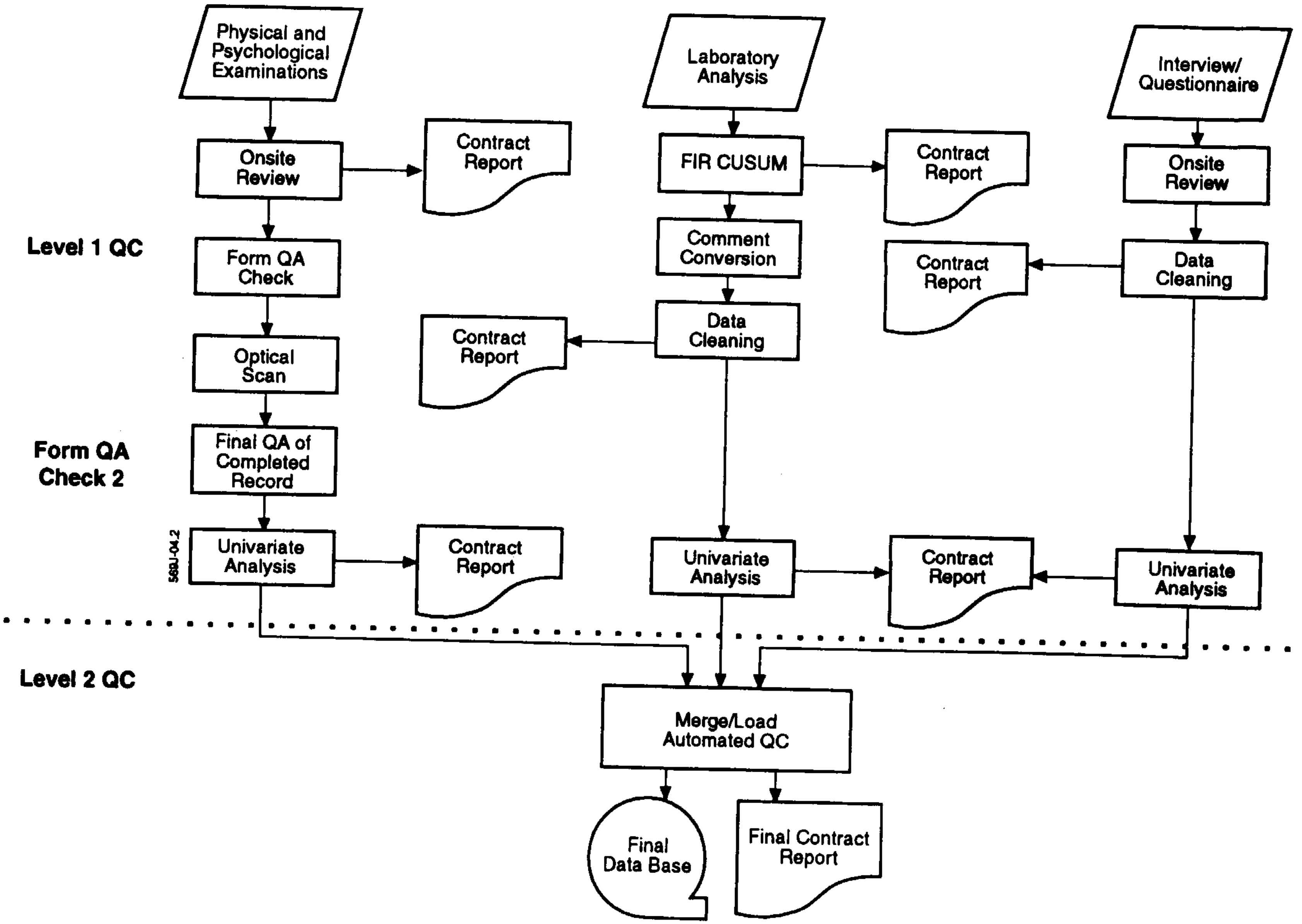

Figure 6-1.
Two Levels of Quality Control Applied to All Collected Data Prior to Statistical Analysis

Design and Administration of Physical and Psychological Examination Forms

As mentioned previously, the examination forms were designed to solicit all required data such that recording time was minimized, comprehension was enhanced, and data input could occur with a minimum of transcription errors. Optical mark recognition (OMR) technologies were selected to eliminate the risk of transcription errors and were applied to all psychological tests. Customized mark-sense forms also were developed and OMR technology was used to achieve these same objectives for segments of the physical examination and the self-administered questionnaires. The use of mark-sense forms allowed the creation of computerized data files directly from the raw data recorded on these forms.

QC procedures for all data collection instruments began with a review of each form as it was completed. A mark-sense reader was used at SCRF to scan for completeness and to conduct some broad-based logic checks. Any forms containing missing, incomplete, or contradictory examination results were returned to the examining physician for completion before the participants left the site. Any questionable results or "hard-to-diagnose" conditions (such as heart sounds or peripheral pulses) were verified by the diagnostician at the outbriefing. In addition, any differences in interpretation between examiners were identified, and adjustments in recording protocols and programmed data extraction were made as necessary. All examination forms were signed by the examining physician, and the examiner identification number was coded in the data base. A final level of QC audit was accomplished by Air Force statisticians, who conducted a detailed screening of the data and checked for errors.

Data Completeness Checks

Customized programming of the OMR allowed for the identification of those forms (and their corresponding data records) with missing responses, as well as those with multiple responses to questions that required a single response. The OMR scanner was programmed to reject forms that failed completeness and multiple response checks and generate control code for each rejected form. The control code identified the location of all verification checks failed for a given form.

When a raw data form was rejected, the reason for the rejection was determined and the exact data element corrected by comparing the rejected raw data form to the values recorded in the data record created by the scanner. A customized set of rejection and resolution codes was developed for the study to describe all the reasons for a form's rejection and any subsequent reasons for changing a data value. Various codes identified values recovered from light marks, missing marks explained by examiner comments, and missing comment flags resolved by the presence or absence of text in the comment areas. These codes ensured data completeness by accounting for all questionable or missing responses.

Some of the rejected forms did not contain actual data errors, but rather, anomalies created in using mark-sense cards for data collection. For example, the scanner incorrectly counted incompletely erased responses, and missed responses marked with too little carbon or graphite. Also, examiners tended to mark responses clearly for abnormal findings and to mark responses lightly or to bypass responses for expected or desired findings. Failure of the form to provide the correct number of expected responses always resulted in rejection. These errors were resolved, as were the anticipated, more traditional errors.

The rejection code, data location code, resolution code, data inspector's initials, and correct data value were posted directly on a participant's data record. This procedure not only effectively maintained a comprehensive audit trail of all record manipulations, but also provided a mechanism for measuring the frequency of specific errors.

Statistics were compiled on out-of-range results and data omissions that had been accepted in the previous QC audits. The results were monitored to detect trends, possible bias situations, and other data quality problems. This information was reviewed and relayed to examiners and internal auditors to assist in preventing or correcting chronic, but avoidable, problems. Refresher training was provided to examining physicians to avoid data omissions. Physicians were consulted to recover missing data, and out-of-range results were reviewed for logical validity by an independent clinician.

Data Validation

Data files were examined in a series of verification and validation procedures developed to check the results within each participant's record for logical consistency and abnormal findings. Any records noted to have ambiguous findings, incongruent observations, extreme results, errors, or omissions were listed and submitted for review to a physician.

Again, clinical judgments were made by the auditing physician in assigning a validation code for each extreme or questionable data result. The validation codes allowed the physicians to indicate that data were deciphered from examiner comments or from related findings from another specialty area, or were accurately recorded and logically consistent with other findings for the participant. Data items that could not be definitively validated or recovered through clinical judgment and consultation with the original examiner were assigned codes noting missing or invalid data values. Some reasons for unavailable data included the following:

- Participant refusal
- Incomplete, confusing, ambiguous, or unclassifiable information
- Contaminated samples
- Unscorable psychological examinations
- Use of data from previous Air Force studies, at which the 1992 participant was not present
- Exemption from testing (e.g., exemption from delayed skin testing to prevent confounding of immunology panel results).

These unrecoverable data were excluded from subsequent analysis. The number of values not available for analyses is presented in the clinical chapters by variable.

Medical Records Coding Quality Control

SAIC forwarded completed questionnaires and physical examination records to the Air Force at Brooks Air Force Base, Texas, for diagnostic coding and verification of all subjectively reported conditions. The Air Force used the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) for morbidity coding; the Systematized Nomenclature of Medicine for anatomic site coding; and the American Hospital Formulary Service for medication coding. Two coders independently processed each questionnaire and physical examination. Both codings were then subjected to a 100-percent QA and QC review, during which every posted code was checked against medical records. A third coder adjudicated any discrepancies.

After QA and QC review and/or adjudication, information from the coding sheets was placed into the AFHS data base using 100-percent, double-blind data entry and verification. Any discrepancies were reviewed, corrected, and again subjected to double-blind data entry and verification. After coding and data entry, the Air Force batched the questionnaires and forwarded them to NORC in Chicago, Illinois, for data processing. The Air Force then obtained the NORC questionnaire data tape, matched this information to the Air Force data file, and resolved any differences. A single, final combined data base was produced by NORC, and a copy sent to the Air Force.

Processing of Questionnaire Data

All questionnaires completed at the examination site were edited twice: first by the interviewer who administered the questionnaire, and then by the site supervisor or editor. These reviews were conducted prior to each participant's departure from the examination site, so that any missing information could be retrieved from the participant onsite. Completed questionnaires, with the exception of the Diet Assessment Questionnaire from the Interval interview, were sent to the Air Force for medical coding. Diet Assessment Questionnaires were sent directly from the examination site to the NORC Chicago office.

After completion of the medical coding, questionnaires were sent to the NORC Chicago office for data processing. Upon receipt, questionnaires were logged into the receipt control system. Diet Assessment Questionnaires were sent to the subcontractor, Willett Associates, for coding and data entry. The rest of the questionnaires were processed by NORC in Chicago.

To process the questionnaires, NORC first coded responses to open-ended questions and key entered the data into a data base. Data entry was 100-percent verified by a second key entry operator. Then, an editing program was executed that checked for valid value ranges, inter-item consistency, and correct logic, dates, and arithmetic. The editing program produced an error sheet for each questionnaire in which a discrepancy was identified. Questionnaires were reviewed to resolve discrepancies on a case-by-case basis. No changes were ever made to the hard copy data; corrections were entered only into the data base and the editing program was re-run. This process was repeated until no errors were detected.

STATISTICAL ANALYSIS QUALITY CONTROL

Specific QC measures were developed for the statistical analysis task efforts, such as construction of data bases for the statistical analysis of each clinical chapter, the statistical analysis itself, and the preparation of the clinical chapters.

Each specialized statistical data base was constructed by defining and locating every variable within the many subparts of the composite followup data base. Although the data had been subjected to QC procedures during collection and were frozen prior to starting the statistical analysis, statistical checks for outliers and other improbable values were conducted; anomalies identified by the statisticians were discussed with those responsible for the data collection (i.e., NORC, SCRF, or the Air Force).

QA largely depended on regular communication and general agreement among statisticians. Several meetings and consultations between the Air Force team and SAIC statisticians were held in conjunction with the development of the data analysis plan. Additionally, frequent telephone conversations took place during the course of the physical exam. During the analysis, there were frequent telephone conversations and any problems identified in the statistical analysis were resolved by team discussion. The software was checked by comparing results from analyses on the same variable by different programs. The statisticians frequently checked to determine that the number of observations used in an analysis was correct, and peer review ensured that the program code was appropriate for the chosen procedure. The analyses were conducted in accordance with the data analysis plan, which was reviewed extensively. Throughout the study, the Air Force and SAIC maintained duplicate data bases. Upon completion of the analyses, SAIC delivered all analysis software and SAS® data sets for each clinical area to the Air Force for final review and archiving.

All tables and statistical results were checked against the computer output from which they were derived, and all statistical statements in the texts were checked for consistency with the results given in the tables. In addition, drafts of each chapter in this report were reviewed by the Air Force and SAIC investigators and the SAIC Quality Review Committee (QRC).

Data Base Modifications

After the statistical analyses were underway, errors were discovered in the data base. One participant was coded in the data base as Black, when he was actually non-Black. After the data base had been created, one additional Ranch Hand was found to have a history of hepatitis C. Also, due to discrepancies in the heights coded in the data base, body fat measurements were incorrect for 17 participants.

The non-Black participant who was coded as Black in the data base was a 50-year-old Comparison in the enlisted flyer cohort with a current serum dioxin value less than 10 ppt. Because he is a Comparison, he was only included in the Model 1 and Model 3 analyses (see

Chapter 7, Statistical Methods). Race was used as a candidate covariate in the analyses of all clinical chapters; therefore, this Comparison was erroneously used as a Black, rather than as a non-Black participant in the Model 1 and Model 3 analyses of all clinical chapters. In the Neoplasia Assessment, this participant was excluded from the analyses of melanoma because he was erroneously coded as Black. However, additional analyses of melanoma were performed with the participant properly coded as non-Black and the conclusions from the analyses did not change in response to this misclassification.

The data base was corrected to account for the Ranch Hand that was found to have a history of hepatitis C after the data base had been created. However, statistical analyses in the Gastrointestinal Assessment that excluded participants with a presence of hepatitis C antibodies were underway prior to discovery of this misclassification. This Ranch Hand did not have a dioxin measurement and therefore only the results of Model 1 were affected. The corrected data base was used for the statistical analyses of the dependent variable "Antibodies for Hepatitis C."

Body fat measurements in the original data base contained inconsistencies due to variations in participants' heights across examination cycles. Body fat was calculated according to the formula:

$$\text{Body fat (in percent)} = \frac{\text{Weight (kg)}}{[\text{Height (m)}]^2} \times 1.264 - 13.305.$$

For 85 participants, recorded heights fluctuated by more than 5 centimeters across the time of duty in Southeast Asia (SEA) and 1982, 1985, 1987, and 1992 examination records. Discrepancies in recorded heights between the 1992 physical examination form and the 1992 pulmonary examination form were identified for 17 of these participants. Heights recorded in the data base for these 17 participants were replaced with the reported heights from the pulmonary form because these heights were closer to reported heights from previous cycles. However, the data base was corrected after statistical analyses of Model 1 were started for the four clinical areas in which body fat was used as a candidate covariate (General Health Assessment, Cardiovascular Assessment, Endocrine Assessment, and Pulmonary Assessment). Therefore, the revised body fat measurements were used only in the analyses of Models 2 through 6 in these four clinical chapters; Model 1 analyses used the original body fat measurements. In future cycles, additional QC procedures will be implemented to ensure consistent height measurements across and within the examination cycles.

Statistical Longitudinal Analysis Implications Due to a Change in Laboratory Equipment

Some of the chemical determinations analyzed in this study were performed at the 1982 Baseline examination with a Dupont® Automated Chemical Analyzer and at the 1992 followup examination with a Baxter/Dade Paramax® Automated Chemical Analyzer. This change was dictated by new technology and an increase in efficiency with high precision and reliability. However, because longitudinal analyses require contrasts of changes from 1982 to 1992, a concern was raised that the change in equipment might bias the outcome of the longitudinal statistical analyses.

For the eight chemical determinations to be investigated in the statistical longitudinal analysis that were measured with the Dupont® equipment at the Baseline examination and the Baxter/Dade Paramax® equipment at the 1992 followup examination, a comparison of the two instruments was conducted. The eight chemical determinations of interest were aspartate aminotransferase (AST), alanine aminotransferase (ALT), gamma glutamyl transferase (GGT), total cholesterol, high-density lipoprotein (HDL) cholesterol, triglycerides, fasting glucose, and serum creatinine. For each of the eight chemical determinations, the machines were compared on each of 30 split tri-level control samples (10 samples each at low, medium, and high levels) and the results were plotted (Dupont® versus Baxter/Dade Paramax® determination) and statistically assessed for linearity. All analyses exhibited a high degree of linearity (4). Because these chemical determinations were longitudinally analyzed with linear models, these results suggest that the change in instrumentation was not a source of bias.

ADMINISTRATIVE QUALITY ASSURANCE

In recognition of the magnitude, complexity, and importance of the AFHS, the QRC was established by SAIC at the initiation of the 1985 followup and continued through the 1987 and 1992 followup studies for the purpose of providing general oversight to the AFHS program and advice on the appropriateness of program management and QC actions. The QRC was composed of SAIC senior corporate personnel. These independent reviewers remained separate from the project management staff. The QRC met periodically to review recent study progress and any issues that either had an impact on study quality or were perceived as a potential problem. Members of the QRC also conducted first-hand evaluations of ongoing program operations.

CHAPTER 6
REFERENCES

1. Bissell, A.F. 1969. CUSUM techniques for quality control. *Applied Statistics* Vol. 18.

2. Lucas, J.M., and R. Crosier. 1982. Fast initial response for CUSUM quality control schemes: Give your CUSUM a headstart. *Technometrics* Vol. 24.

3. Grubbs, F.E. 1969. Procedures for detecting outlying observations in samples. *Technometrics* XI:1-21.

4. ACA/Paramax Comparison. 14 October 1993. SAIC Document 0312.